Available online at www.sciencedirect.com
SCIENCE DIRECT®
Discrete Optimization 3 (2006) 50–62
DISCRETE OPTIMIZATION
www.elsevier.com/locate/disopt

On the high multiplicity traveling salesman problem

Alexander Grigoriev[a], Joris van de Klundert[b,*]

[a] Department of Quantitative Economics, Faculty of Economics and Business Administration, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands
[b] Department of Mathematics, Faculty of Science, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands

Received 20 September 2005; received in revised form 3 October 2005; accepted 3 November 2005
Available online 24 January 2006

Abstract

This paper considers a version of the traveling salesman problem where the cities are to be visited multiple times. Each city has its own required number of visits. We investigate how the optimal solution and its objective value change when the numbers of visits are increased by a common multiplicator. In addition, we derive lower bounds on values of the multiplicator beyond which further increase does not improve the average tour length. Moreover, we show how and when the structure of an optimal tour length can be derived from tours with smaller multiplicities.

Keywords: Traveling salesman problem; High multiplicity

1. Introduction

In this work, we consider a generalization of the traveling salesman problem where, given a complete undirected graph with vertex set $J = \{1, 2, \ldots, n\}$, a distance matrix $[c_{i,j}]_{n\times n}$, and positive integers $s_i, i \in J$, we have to find a minimal length closed tour in which vertex $i \in J$ is visited exactly s_i times. We assume that all elements of matrix $[c_{i,j}]_{n\times n}$ are non-negative integers. In particular we assume that, in contrast to the classical traveling salesman problem, elements $c_{i,i}$ are not necessarily zeros. Moreover, in the graph as well as in the tour we do allow loops. We also assume that all parameters in the input are finite. This problem has been studied under the name of traveling salesman problem with many visits to few cities; see, e.g., [4].

This paper considers a problem version which is more general than previously studied versions, and we refer to it as the High Multiplicity Traveling Salesman Problem (HMTSP). The HMTSP, which is of interest in itself, also serves as a basic high multiplicity sequencing problem. Recently, there has been considerable interest in high multiplicity scheduling and sequencing problems [1,3,7,8], most of which is due to their applicability in current manufacturing and logistics management. For instance, no-wait flow shop scheduling problems in which there is a multitude of jobs with identical processing requirements have been widely studied (see e.g. [1]). Such problems arise for example in the context of assembly line optimization. Another application is in the field of scheduling airplane

* Corresponding author.
E-mail addresses: a.grigoriev@ke.unimaas.nl (A. Grigoriev), j.vandeklundert@math.unimaas.nl (J. van de Klundert).

1572-5286/$ - see front matter
doi:10.1016/j.disopt.2005.11.002

maintenance at airports [4]. All these problems can be naturally formulated as an HMTSP. Therefore, apart from its combinatorial interest, an improved understanding of the problem contributes to an improved understanding of more applied problems in, for instance, flow shop scheduling, repetitive manufacturing and cyclic scheduling.

The program of this paper is as follows. First of all, we mathematically define the problems using a well known integer linear programming (ILP) formulation, and we discuss how to encode solutions of the problem. The point of interest here is that encoding a solution by explicitly giving the sequence of visits is not a polynomial encoding of the output (when assuming a 'reasonable' encoding of the input; see [6]). Nevertheless, the problem is easily seen to be in NP and we present encoding schemes for the output and results on the structure of optimal solutions.

Subsequently, we set out to study how the objective value F of the ILP formulation and the structure of the solution change when the numbers of visits to the vertices are increased by a common multiplicator ℓ. We are interested in the asymptotic behavior of the function $F(\ell)$ defined as the average value of the objective function when the number of visits to the vertices is increased by multiplicator ℓ. It will become clear that for some instances $F(\ell)$ is a strictly decreasing function, whose asymptotic value equals the objective value F_T of a relaxation of the ILP formulation (which results after deleting the subtour elimination constraints). For other instances, there exists ℓ^0, such that for $\ell \geq \ell^0$, $F(\ell)$ equals F_T. Such instances are called stable. We show that deciding whether an instance is stable can be done in polynomial time. We also present an upper bound on ℓ^0. This bound is linear in n, and is shown to be tight. In fact, we will show a much stronger result, namely that for $\ell \geq n-1$, $(\ell+1)F(\ell+1) = \ell F(\ell) + F_T$, for any instance, stable or not.

2. On the complexity of the problem

First of all, let us show that the decision version of the HMTSP is in NP. This follows from the fact that the HMTSP can be formulated by the following generalization of a well known integer program formulation for the classical traveling salesman problem; see [13]:

$$\min_x \sum_{i\in J}\sum_{j\in J} c_{i,j}x_{i,j} \tag{1}$$

subject to

$$\sum_{i\in J} x_{i,j} = s_j, \quad j \in J; \tag{2}$$

$$\sum_{j\in J} x_{i,j} = s_i, \quad i \in J; \tag{3}$$

$$\sum_{i\in J'}\sum_{j\in J\setminus J'} x_{i,j} \geq 1, \quad \forall J' \subset J \text{ such that } J' \neq \emptyset; \tag{4}$$

$$x_{i,j} \in \mathbb{Z}^+, \quad i \in J,\ j \in J, \tag{5}$$

where $x_{i,j}$ indicates the number of times the high multiplicity traveling salesman tour uses arc (i, j) leaving vertex i and entering vertex j. Recall that, since loops are allowed, it is possible that $x_{i,i} > 0$.

For a generic instance of (1)–(5) we will refer to this formulation as P, and to the value of its optimal solution as $F(P)$. Notice that the HMTSP differs from the classical TSP only by the right-hand side of the constraints (2) and (3): in the classical TSP all right-hand-side coefficients are equal to one.

Despite the exponential number of constraints, a solution of the program forms a polynomially encoded certificate, which can be checked in polynomial time, as required for membership in NP. This follows from the fact that for any integral solution of P, the *subtour elimination constraints* (4) are satisfied if and only if the set of arcs $\{(i, j) : x_{i,j} \geq 1\}$ forms a strongly connected digraph on J, and strong connectivity of a graph can be checked in polynomial time. Moreover, using Lenstra's algorithm for ILP in fixed dimension, see [11], the HMTSP can be solved in polynomial time when the number of vertices in G is fixed. In fact, combinatorial, strongly polynomial, algorithms for the case where the number of vertices is bounded from above by a constant have been developed by Cosmadakis and Papadimitriou, Kanelakis and Papadimitriou, and Van de Klundert in [4,9,10].

Nevertheless, the following question remains: how to construct a tour (the sequence of the vertices), given a solution of P. Here, we briefly explain how an (optimal) *sequence* can be constructed from an (optimal) solution x^0 to P

(see, for instance, [5,10,14]). Let $[x^0_{i,j}]_{n\times n}$ be an (optimal) solution to P. We are going to describe a procedure *ConvertToSequence* that creates a directed multigraph on node set J such that an Eulerian trail in the multigraph will correspond to the optimal traveling salesman tour that traverses exactly $x^0_{i,j}$ copies of arc (i, j), $i \in J, j \in J$. Let us recall that an Eulerian trail is a closed walk in a directed multigraph that visits each arc of the multigraph exactly once.

We denote by $G(J, A)$ the multigraph where the nodes correspond to the vertices in the HMTSP, and each arc (i, j) occurs with multiplicity $x^0_{i,j}$. In the procedure *ConvertToSequence* we use the phrase 'simple closed walk', which is to be understood as a simple directed cycle in G. More precisely, we define a directed cycle $C = ((i_1, i_2), (i_2, i_3), \dots, (i_t, i_1))$, $i_\tau \in J$, $\tau \in \{1, 2, \dots, t\}$ to be a *simple closed walk* if it does not contain any closed subwalk, i.e., if $i_{\tau'} \neq i_{\tau''}$ for any two different indices $\tau', \tau'' \in \{1, 2, \dots, t\}$.

Procedure ConvertToSequence:

Input: $[x^0_{i,j}]_{n\times n}$.

Output: A closed walk in the graph $G(J, A)$, represented by a finite collection $\mathcal{C}$ of pairs (m_C, C), where C is a simple closed walk and m_C is an integer corresponding to the number of copies of walk C in the collection $\mathcal{C}$.

Algorithm:

1. $\mathcal{C} := \emptyset, q = 1$.
2. Find a simple closed walk $C_q = ((i_1, i_2), (i_2, i_3), \dots, (i_t, i_1))$, $i_\tau \in J$, $\tau \in 1, 2, \dots, t$, such that $x^{q-1}_{i_\tau, i_{\tau+1}} > 0$, $\tau \in \{1, 2, \dots, t-1\}$ and $x^{q-1}_{i_t, i_1} > 0$.
3. Let $m_q = \min\{x^{q-1}_{i,j} : (i, j) \in C_q\}$, $\mathcal{C} := \mathcal{C} \cup \{(m_q, C_q)\}$ and $x^q_{i,j} := x^{q-1}_{i,j} - m_q$ for $(i, j) \in C_q$ and $x^q_{i,j} := x^{q-1}_{i,j}$ otherwise.
4. If $[x^q_{i,j}]_{n\times n} = [0]_{n\times n}$, output: $\mathcal{C}$ and stop, else set $q := q + 1$, and goto step 2.

By the integer program formulation, the directed multigraph formed by the union of the cycles in $\mathcal{C}$ is strongly connected. Since $\mathcal{C}$ is a collection of closed walks, all in-degrees of nodes equal to out-degrees. This implies that there exists an Eulerian trail in G. Let W be such an Eulerian trail. Polynomial time algorithms for finding an Eulerian trail in an Eulerian multigraph are well known and can be found in several text books; see, e.g., [12]. By construction, for any $i \in J$, trail W visits i exactly s_i times. Therefore, the traveling salesman tour corresponding to W is feasible and optimal for P. For the correctness of the procedure *ConvertToSequence* we also refer to [5].

Let us briefly discuss the time complexity of the procedure. First of all, it is not hard to verify that step 2 can be executed in $O(n^2)$ time. Further, by definition of step 3, $[x^q_{i,j}]_{n\times n}$ contains at least one more zero element than $[x^{q-1}_{i,j}]_{n\times n}$, hence the algorithm terminates after $O(n^2)$ iterations. Steps 3 to 4 are executed in at most $O(n)$ time. This leaves the overall time complexity of *ConvertToSequence* to be $O(n^4)$. In addition, this reasoning implies that the collection $\mathcal{C}$ consists of at most $O(n^2)$ pairs (m_C, C). Since m_C and C are also polynomially bounded in the input size, the output of *ConvertToSequence* encodes an optimal solution polynomially in the input size. Hence, we have two compact encoding schemes of optimal solutions: $\mathcal{C}$, and $[x^0_{i,j}]_{n\times n}$. Subsequent sections will mostly use the integer linear programming solution $[x^0_{i,j}]_{n\times n}$.

3. Definitions

We now continue by considering some relaxations and extensions of the HMTSP. For a generic instance we define T to be the problem that results from P by relaxing the subtour elimination constraints (4).

Definition 1 (*Transportation Problem T*).

$$F_T = \min_x \sum_{i\in J} \sum_{j\in J} c_{i,j} x_{i,j} \tag{6}$$

subject to

$$\sum_{i\in J} x_{i,j} = s_j, \quad j \in J; \tag{7}$$

$$\sum_{j\in J} x_{i,j} = s_i, \quad i \in J; \tag{8}$$

$$x_{i,j} \in \mathbb{Z}^+, \quad i \in J,\ j \in J. \tag{9}$$

We denote the value of the optimal solution of T by F_T. Problem T is a transportation problem, which satisfies the following property.

Property 1 (*Integrality of the Transportation Polytope*). *The polytope defined by (7) and (8) is integral, i.e., all its vertices are defined by integral vectors* [13].

As a well known consequence of this property, T always has an integer optimal solution, which can be found in polynomial time.

Finally, we introduce the more general problem of minimizing the average tour length. To this purpose, we propose a formulation that allows more general tour structures where in the resulting tour the number of node visits is multiplied by a factor of $\ell \in \mathbb{N}$.

Definition 2 (*Parametrized HMTSP P_ℓ*). Given a positive integer parameter $\ell \in \mathbb{N}$, P_ℓ is defined as

$$F(\ell) = \frac{1}{\ell} \min_x \sum_{i\in J}\sum_{j\in J} c_{i,j}x_{i,j} \tag{10}$$

subject to

$$\sum_{i\in J} x_{i,j} = \ell s_j, \quad j \in J; \tag{11}$$

$$\sum_{j\in J} x_{i,j} = \ell s_i, \quad i \in J; \tag{12}$$

$$\sum_{i\in J'}\sum_{j\in J\setminus J'} x_{i,j} \geq 1, \quad \forall J' \subset J \text{ such that } J' \neq \emptyset; \tag{13}$$

$$x_{i,j} \in \mathbb{Z}^+, \quad i \in J, j \in J. \tag{14}$$

Notice again that P_ℓ can be solved in polynomial time when n is fixed [11]. We refer to the objective function of P_ℓ by $F(P_\ell)$, or $F(\ell)$ for short.

Now, the problem is to find a parameter ℓ^* that minimizes $F(\ell)$ over all $\ell \in \mathbb{N}$. Indeed, given an optimal pair (ℓ^*, x^*), it encodes a sequence in which each node $j \in J$ is visited $\sum_{i\in J} x^*_{i,j} = \ell^* s_j$ times. Let us recall that the aim is to minimize the average tour length. This is achievable by minimizing the minimum tour length over all natural numbers ℓ. Thus allowing ℓ^* to be infinite, the optimal solution value $F(\ell^*)$ specifies the minimum average tour length.

As before, we define T_ℓ to be the problem that results from P_ℓ after relaxing the subtour elimination constraints.

Definition 3 (*Parametrized Transportation Problem T_ℓ*). Given a positive integer parameter $\ell \in \mathbb{N}$, T_ℓ is defined as

$$F_{T_\ell} = \frac{1}{\ell} \min_x \sum_{i\in J}\sum_{j\in J} c_{i,j}x_{i,j} \tag{15}$$

subject to

$$\sum_{i\in J} x_{i,j} = \ell s_j, \quad j \in J; \tag{16}$$

$$\sum_{j\in J} x_{i,j} = \ell s_i, \quad i \in J; \tag{17}$$

$$x_{i,j} \in \mathbb{Z}^+, \quad i \in J,\ j \in J. \tag{18}$$

Notice that defining $y_{i,j} = x_{i,j}/\ell$, the problem T_ℓ can be rewritten as

$$F_{T_\ell} = \min_y \sum_{i\in J}\sum_{j\in J} c_{i,j} y_{i,j} \tag{19}$$

subject to

$$\sum_{i\in J} y_{i,j} = s_j, \quad j \in J; \tag{20}$$

$$\sum_{j\in J} y_{i,j} = s_i, \quad i \in J; \tag{21}$$

$$y_{i,j} \in \mathbb{Z}^+, \quad i \in J,\ j \in J. \tag{22}$$

Hence, we conclude that T_ℓ and T are identical transportation problems. By consequence, for any natural number ℓ the value F_T yields a lower bound on $F(\ell)$.

4. General properties of optimal solutions

In this section we derive some basic properties of the parametrized problem P_ℓ. We derive some results on how the optimal value $F(\ell)$ decreases when ℓ increases. All of these results are based on the following inequality.

Theorem 1. *For any natural number ℓ, the following inequality holds:*

$$F(\ell+1) \le \frac{\ell}{\ell+1} F(\ell) + \frac{1}{\ell+1} F_T. \tag{23}$$

Proof. Let $x^{\ell+1}$ and x^ℓ, $\ell \in \mathbb{N}$, denote optimal solutions for $P_{\ell+1}$ and P_ℓ respectively, and let x^T denote an optimal solution for transportation problem T. Now, notice that matrix $x^\ell + x^T$ is a feasible solution for $P_{\ell+1}$. Thus, for any number $\ell \in \mathbb{N}$ we have

$$\begin{aligned} F(\ell+1) &= \frac{1}{\ell+1}\sum_{i\in J}\sum_{j\in J} c_{i,j} x_{i,j}^{\ell+1} \\ &\le \frac{1}{\ell+1}\sum_{i\in J}\sum_{j\in J} c_{i,j} x_{i,j}^{\ell} + \frac{1}{\ell+1}\sum_{i\in J}\sum_{j\in J} c_{i,j} x_{i,j}^{T} \\ &= \frac{\ell}{\ell+1}\left(\frac{1}{\ell}\sum_{i\in J}\sum_{j\in J} c_{i,j} x_{i,j}^{\ell}\right) + \frac{1}{\ell+1}\sum_{i\in J}\sum_{j\in J} c_{i,j} x_{i,j}^{T} \\ &= \frac{\ell}{\ell+1} F(\ell) + \frac{1}{\ell+1} F_T \end{aligned}$$

as required. □

Let us denote the right-hand side of (23) by $H(\ell+1)$. Clearly, $H(\ell+1)$ is a weighted average of F_T and $F(\ell)$. From the fact that for all $\ell \in \mathbb{N}$ it holds that $F_T \le F(\ell)$ we derive that $H(\ell+1) \le F(\ell)$. By (23) $F(\ell+1) \le H(\ell+1)$, and we immediately have the following corollary.

Corollary 1. *$F(\ell) \ge F(\ell+1)$ holds for any $\ell \in \mathbb{N}$.* □

Moreover, since $H(\ell+1)$ is a weighted average of F_T and $F(\ell)$ the inequalities in $F_T \le H(\ell+1) \le F(\ell)$ are strict if and only if $F_T < F(\ell)$. From this observation and Corollary 1 we have the following theorem.

Theorem 2. *For any $\ell \in \mathbb{N}$ it holds that $F(\ell) > F(\ell+1)$ if and only if $F(\ell) > F_T$.* □

By consequence, we also have the following corollary.

Corollary 2. *If there exists ℓ^0 such that $F(\ell^0) = F_T$ then $F(\ell) = F_T$ for all $\ell \geq \ell^0$.*

Hence, the question arises whether for a given instance there exists ℓ^0 such that $F(\ell^0) = F_T$. Moreover, in the case that the answer is affirmative, one would like to efficiently compute or at least estimate the smallest number ℓ^0 such that equality holds. These questions will be discussed and answered in Sections 5 and 6. In the case that the answer is negative, there might still be some long run stabilization worthy of characterization, which is the subject of Section 7. Pursuing this direction, we first consider the following theorem, which is an extension of Theorem 1.

Theorem 3. *For any fixed natural number ℓ^* and any natural number $\ell \geq \ell^*$ the following inequality holds:*

$$F(\ell) \leq \frac{\ell^*}{\ell} F(\ell^*) + \frac{\ell - \ell^*}{\ell} F_T. \tag{24}$$

Proof. We shall prove the theorem by induction on ℓ from the basis ℓ^*. For any natural number ℓ^* we derive that $F(\ell^*) \leq \ell^* F(\ell^*)/\ell^* + (\ell^* - \ell^*)F_T/\ell^* = F(\ell^*)$, and hence (24) holds for $\ell = \ell^*$.

Now, we prove that (24) holds for $\ell + 1$. From Theorem 1 we obtain that $F(\ell+1) \leq \ell F(\ell)/(\ell+1) + F_T/(\ell+1)$ holds for every natural number ℓ. Hence, by the induction hypothesis we have that

$$\begin{aligned} F(\ell+1) &\leq \frac{\ell}{\ell+1} F(\ell) + \frac{1}{\ell+1} F_T \\ &\leq \frac{\ell}{\ell+1}\left(\frac{\ell^*}{\ell} F(\ell^*) + \frac{\ell-\ell^*}{\ell} F_T\right) + \frac{1}{\ell+1} F_T \\ &= \frac{\ell^*}{\ell+1} F(\ell^*) + \frac{\ell+1-\ell^*}{\ell+1} F_T \end{aligned}$$

and the proof is complete. □

Finally, we prove that the value $F(\ell)$ converges to F_T when ℓ goes to infinity.

Theorem 4.

$$\lim_{\ell \to +\infty} F(\ell) = F_T. \tag{25}$$

Proof. Using Theorem 3, and letting $\ell > \ell^*$ we derive

$$\begin{aligned} F_T &\leq F(\ell) \\ &\leq \frac{\ell^*}{\ell} F(\ell^*) + \frac{\ell-\ell^*}{\ell} F_T \\ &= F_T + \frac{\ell^*}{\ell}\left(F(\ell^*) - F_T\right). \end{aligned}$$

Hence,

$$\begin{aligned} F_T &\leq \lim_{\ell \to +\infty} F(\ell) \\ &\leq \lim_{\ell \to +\infty} \left(F_T + \frac{\ell^*}{\ell}\left(F(\ell^*) - F_T\right)\right) \\ &= F_T. \end{aligned}$$

which yields the desired result. □

We finish this section by demonstrating how optimal solutions and objective function values change with ℓ.

Example 1. Consider the following instance.
Let $J = \{1, 2, 3\}$;

$$\begin{aligned} &s_1 = 1, \quad s_2 = 1, \quad s_3 = 1; \\ &c_{1,3} = c_{3,1} = a, \quad (a > 1); \\ &c_{i,j} = 1 \quad \text{for } (i, j) \notin \{(1, 3), (3, 1)\}. \end{aligned}$$

The reader is encouraged to verify that:

- For the transportation problem T it holds that $F_T = 3$, $x^T_{i,j} = 0$ for $i \neq j$ and $x^T_{i,j} = 1$ for $i = j$.
- For the problem P_1 we have $F(P) = a + 2$. An optimal solution for this problem is $x^1_{i,j} = 1$ for $(i, j) \in \{(1, 2), (2, 3), (3, 1)\}$ and $x^1_{i,j} = 0$ for all other arcs. In this case collection $\mathcal{C}$ as output by *ConvertToSequence* is $\{(1, ((1, 2), (2, 3), (3, 1)))\}$.
- For any natural number $\ell \geq 2$ we get $F(\ell) = 3$. An optimal solution is specified by $x^\ell_{1,2} = x^\ell_{2,3} = x^\ell_{3,2} = x^\ell_{2,1} = 1$; $x^\ell_{1,1} = x^\ell_{3,3} = \ell - 1$; $x^\ell_{3,1} = x^\ell_{1,3} = 0$ and $x^\ell_{2,2} = \ell - 2$. Further,
$$\mathcal{C} = \{((1, (1, 2)(2, 3), (3, 2), (2, 1)), ((\ell - 1), (1, 1)), ((\ell - 2), (2, 2)), ((\ell - 1), (3, 3)))\}.$$

In this example it holds that $F(\ell) = F_T$, for $\ell \geq 2$. This example also demonstrates that the ratio between $F(P)$ and $F(\ell)$ can be arbitrarily large, namely $F(P)/F(\ell) = (a + 2)/3$, where a may be chosen arbitrarily.

Example 2. Consider another instance which is identical to the one described in Example 1, but for the following arc lengths: $c_{1,2} = c_{2,1} = b$, where $1 < b < a$.

It can be checked that in all cases discussed in Example 1 the same solutions are the optimal ones, but the objective function values are different. More specifically:

- For the transportation problem T it holds that $F_T = 3$, $x^T_{i,j} = 0$ for $i \neq j$ and $x^T_{i,j} = 1$ for $i = j$.
- For the problem P_1 we have $F(P) = a + b + 1$. An optimal solution for this problem is $x^1_{i,j} = 1$ for $(i, j) \in \{(1, 2), (2, 3), (3, 1)\}$ and $x^1_{i,j} = 0$ for any other arcs. In this case $\mathcal{C}$ is $\{(1, ((1, 2), (2, 3), (3, 1)))\}$.
- For any natural number $\ell \geq 2$ we get $F(\ell) = (3\ell + 2b - 2)/\ell = 3 + (2b - 2)/\ell$. An optimal solution is again specified by $x^\ell_{1,2} = x^\ell_{2,3} = x^\ell_{3,2} = x^\ell_{2,1} = 1$; $x^\ell_{1,1} = x^\ell_{3,3} = \ell - 1$; $x^\ell_{3,1} = x^\ell_{1,3} = 0$ and $x^\ell_{2,2} = \ell - 2$. Further, $\mathcal{C} = \{((1, (1, 2)(2, 3), (3, 2), (2, 1)), ((\ell - 1), (1, 1)), ((\ell - 2), (2, 2)), ((\ell - 1), (3, 3)))\}$.

In this example, $F(\ell)$ strictly decreases when ℓ increases, and indeed there does not exist ℓ^0 such that $F(\ell^0) = F_T$. Nevertheless, the optimal solution displays the same (stable) behavior as in the previous example.

5. Stabilization recognition

In this section we study the behavior of instances for which there exists $\ell \in \mathbb{N}$ such that the parametrized problem P_ℓ has solution value $F(\ell) = F_T$, the lower bound which denotes the optimal solution value of the transportation problem T. Such instances will be called stable and we address the computational complexity of stable instance recognition.

Definition 4 (*Stable Instance*). We call an instance of the HMTSP stable if there exists a number $\ell \in \mathbb{N}$ such that $F(\ell) = F_T$.

The following lemma will be useful to prove the main result of this section and in subsequent sections.

Lemma 1. *For every stable instance of the HMTSP, and any proper partition $I, \overline{I}$ of set J ($I \neq \emptyset$, $I \subset J$, and $\overline{I} = J \setminus I$), there exists an integer optimal solution z to the underlying transportation problem T such that $z_{i^*,j^*} > 0$ for some $i^* \in I$ and $j^* \in \overline{I}$.*

Proof. Since the instance is stable, there exists a number ℓ such that $F(\ell) = F_T$. Let x^ℓ be an optimal solution to P_ℓ. Define $y^\ell_{i,j} = x^\ell_{i,j}/\ell$, $(i, j) \in J \times J$. Note that y^ℓ is an optimal solution to T. Notice also that x^ℓ is a feasible solution to P_ℓ and hence satisfies the subtour elimination constraints. Thus, for any proper partition $I, \overline{I}$ of set J, there exists a pair (i^*, j^*) such that $i^* \in I$, $j^* \in \overline{I}$, and $y^\ell_{i^*,j^*} > 0$. Because of Property 1, the optimal solution y^ℓ to transportation problem T is a convex combination of optimal integral solutions to T. Therefore, there must exist an optimal integer solution z to T with strictly positive z_{i^*,j^*}. $\square$

Definition 5 (*Complete Walk*). Given a directed multigraph $G = (J, A)$ and a closed walk $W = ((j_1, j_2), (j_2, j_3), \ldots, (j_w, j_1))$, $j_\tau \in J, \tau \in \{1, 2, \ldots, w\}$ in G. We say W is complete if it visits every node in G, i.e., for every $j \in J$ there exists $\tau \in \{1, 2, \ldots, w\}$ such that $j_\tau = j$.

The main result of this section is formulated in the following theorem.

Theorem 5. *The problem of deciding whether an instance is stable can be solved in polynomial time.*

Proof. The proof works as follows. Given an instance of HMTSP, we construct a multigraph G with the property that the instance is stable if and only if G is strongly connected. Multigraph G will be constructed in polynomial time. Since it can be checked in polynomial time whether a multigraph is strongly connected (see, e.g., [2]), the theorem then follows.

Firstly, we are going to check for each arc $(p, q) \in J \times J$ whether there exists an optimal solution of the transportation problem T that contains this arc. This can be achieved by solving a modification of T, where the modification consists of decreasing s_p by one in the outflow constraint (3) for node p, and decreasing s_q by one in the inflow constraint (2) for node q:

$$F_{p,q} = \min_x \sum_{i \in J} \sum_{j \in J} c_{i,j} x_{i,j} \tag{26}$$

subject to

$$\sum_{i \in J} x_{i,j} = s_j, \quad j \in J \setminus \{q\}; \tag{27}$$

$$\sum_{j \in J} x_{i,j} = s_i, \quad i \in J \setminus \{p\}; \tag{28}$$

$$\sum_{i \in J} x_{i,q} = s_q - 1; \tag{29}$$

$$\sum_{j \in J} x_{p,j} = s_p - 1; \tag{30}$$

$$x_{i,j} \in \mathbb{Z}^+, \quad i \in J,\ j \in J. \tag{31}$$

Let us call this modified problem $T_{p,q}$ and denote the optimal objective value by $F_{p,q}$. Obviously, $T_{p,q}$ is again a transportation problem that can be solved in polynomial time. The aforementioned optimal solution of T containing (p, q) exists if and only if $F_T = F_{p,q} + c_{p,q}$. Let Q be the number of arcs (p, q) satisfying $F_T = F_{p,q} + c_{p,q}$.

Now we are ready to construct a directed multigraph $G = (J, A)$. For all (p, q) such that $F_T = F_{p,q} + c_{p,q}$, let $x^{p,q}$ be obtained by slightly modifying an optimal solution x' of $T_{p,q}$. The modification consists in increasing the (p, q)-element by one: $x^{p,q}_{i,j} = x'_{i,j} + 1$ if $(i, j) = (p, q)$ and $x^{p,q}_{i,j} = x'_{i,j}$ otherwise. The arc set A of G is defined as follows. For all (i, j), the multiplicity of arc (i, j) is defined to be the sum of the $x^{p,q}_{i,j}$ over all (p, q) such that $F_T = F_{p,q} + c_{p,q}$.

If G is strongly connected then it contains an Eulerian trail W (see, e.g., [5]), i.e., a complete closed walk visiting every arc in this directed multigraph exactly once. Clearly W visits every node as well as every arc and hence satisfies the subtour elimination constraints. Moreover, W visits node $j \in J$ exactly Qs_j times. Hence, the traveling salesman tour corresponding to W is a solution to P_ℓ with $\ell = Q$.

Conversely, suppose that the instance is stable and G is not strongly connected. Since G is not strongly connected, there exists a proper partition $I, \overline{I}$ of set J such that G does not contain any arc (i, j) for all $i \in I$, $j \in \overline{I}$. On the other hand, since the instance is stable, by Lemma 1 there must exist an optimal integer solution z to T such that $z_{i^*,j^*} > 0$ for some $i^* \in I$ and $j^* \in \overline{I}$. Then, by definition of A, it holds that $i^*, j^* \in A$ and we have arrived at a contradiction. □

6. Stabilization number

Consider the parametrized problem P_ℓ and transportation problem T, with optimal solution values $F(\ell)$ and F_T respectively. If $F(\ell^0) = F_T$ for some $\ell^0 \in \mathbb{N}$, then it must hold that $F(\ell) = F_T$ for all $\ell \geq \ell^0$ because of Corollary 2. Hence, we set out to find the smallest integer ℓ^0 for which $F(\ell^0) = F_T$.

Definition 6 (*Stabilization Number*). For any stable instance of HMTSP, the stabilization number is defined as the smallest integer ℓ^0 for which $F(\ell^0) = F_T$.

Definition 7 (*Solution Induced Graph*). Let x be an integral optimal solution of the transportation problem T. Then G_x is the directed multigraph induced by the set of arcs $A_x = \{(i, j) : x_{i,j} \geq 1\}$ where the arc multiplicity for $(i, j) \in A_x$ is $x_{i,j}$.

Graph G_x is not necessarily strongly connected, since the subtour elimination constraints (4) are relaxed in T. Therefore, we may view the graph G_x as a collection of directed cycles visiting node $i \in J$, exactly s_i times.

Definition 8 (*Components Number*). For any integral optimal solution x of the transportation problem T, U_x is defined to be the number of strongly connected components in G_x. Moreover, $U = \min_x U_x$.

For a pair x, y of integral optimal solutions of the transportation problem T, we let $G_{x+y} = G(J, A_x \cup A_y)$, and we let U_{x+y} be the number of strongly connected components of G_{x+y}. Now, let x and y be such that there exist $i, j \in J$ which are in different strongly connected components of G_x, but arc $(i, j) \in A_y$. Clearly arc (i, j) is in some directed cycle C of G_y. By definition of i and j, the set of vertices on the cycle C intersects with at least two strongly connected components of G_x. Because C is a directed cycle in A_y, the two or more strongly connected components of G_x whose vertex set intersects the vertex set of C are contained in one single strongly connected component in G_{x+y}. This leads to the following observation.

Observation 1. *Let x and y be solutions to T such that there exist $i, j \in J$ with the property that i and j are in different strongly connected components of G_x but $(i, j) \in A_y$. Then $U_{x+y} < U_x$.*

This observation will be used in the proof of the following theorem.

Theorem 6. *For every stable instance, the stabilization number $\ell^0 \leq U \leq n - 1$ and these bounds are tight.*

Proof. We first prove the first inequality. Consider an optimal solution x of the transportation problem T such that the number of strongly connected components in G_x equals U. If $U = 1$, there exists a solution for T satisfying the subtour elimination constraints, and hence the first inequality holds. Hence assume $U > 1$. Let I^1 be a strongly connected component of G_x. Because the instance is stable, Lemma 1 implies that there exists an integer optimal solution z^1 to the underlying transportation problem T such that $z^1_{i^*,j^*} > 0$ for some $i^* \in I^1$ and $j^* \in \overline{I^1}$. By Observation 1 it then holds that the number of strongly connected components of G_{x+z^1} is strictly smaller than the number of strongly connected components of G_x. Now, if $U_{x+z^1} > 1$, let I^2 be one of its strongly connected components. Again by Lemma 1, there exists an integer optimal solution z^2 to the underlying transportation problem T such that $z^2_{i^*,j^*} > 0$ for some $i^* \in I^2$ and $j^* \in \overline{I^2}$. Repeating the same arguments leads to a strongly connected multigraph in at most $U - 1$ steps. Let $k < U$ be this number of steps.

Since $z^1, \ldots, z^k$ are optimal solutions to the transportation problem T, it must hold that $\frac{1}{k+1}(cx + cz^1 + \cdots + cz^k) = cx$. Moreover, $x + z^1 + \cdots + z^k$ is an integral optimal solution to T_{k+1} satisfying the subtour elimination constraints and therefore an optimal solution to P_{k+1}. Combined with Corollary 2 this yields that $F_U = F_{k+1} = F_T$ and hence $\ell^0 \leq U$, as suffices to prove the first inequality.

We continue by showing that in any stable instance $U \leq n - 1$. By definition, $U \leq n$. Hence assume that $U = n$. This implies that T has a single optimal solution consisting solely of arcs (j, j), $j \in J$. Clearly, as can be verified from the proof of Theorem 5, this contradicts the stability of the instance. Hence $U \leq n - 1$.

We complete the proof by showing that the inequalities may hold with equality, i.e., there exists an instance for which $\ell^0 = n - 1$. Consider the following straightforward extension of Example 1.

Example 3. Let $J = \{1, 2, \ldots, n\}$, $s_i = 1$ for all $i \in J$ and the distance matrix is

$$[c_{i,j}]_{n\times n} = \begin{pmatrix} 1 & 1 & 1 & 1 & \ldots & 1 & 1 \\ 1 & 1 & a & a & \ldots & a & a \\ 1 & a & 1 & a & \ldots & a & a \\ \vdots & & & \ddots & & & \\ 1 & a & \ldots & a & 1 & a & a \\ 1 & a & \ldots & a & a & 1 & a \\ 1 & a & \ldots & a & a & a & 1 \end{pmatrix}$$

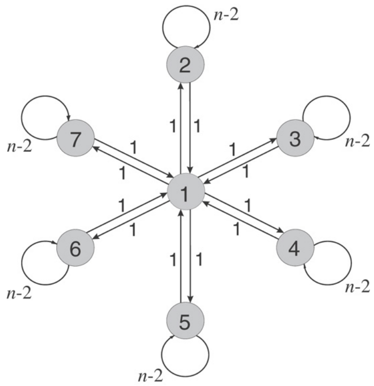

Fig. 1. Optimal tour for $\ell = n - 1$ in Example 3 with $n = 7$.

where $a > 1$.

- For the transportation problem $F_T = n$, $x^T_{i,j} = 0$ for $i \neq j$ and $x^T_{i,j} = 1$ for $i = j$.
- For $\ell = 1$, an optimal solution is given by $x_{i,i+1} = 1$ for $i = 1, 2, \ldots, n-1$, $x_{n,1} = 1$, and $x_{i,j} = 0,\ i, j \in J$ otherwise. Thus, the value of an optimal solution is in this case $F(1) = 1 + (n-2)a + 1 = 2 + (n-2)a$.
- For $\ell = 2$, an optimal solution is given by $x_{1,2} = x_{2,2} = x_{2,1} = 1$, $x_{1,3} = 1$, $x_{i,i+1} = 1$ for $i = 3, 4, \ldots, n-1$, and $x_{n,1} = 1$, $x_{i,i} = 1,\ i = 3, 4, \ldots, n$, and $x_{i,j} = 0,\ i, j \in J$ otherwise. Thus, in this case the value of an optimal solution equals $F(2) = (3 + 1 + (n-3)a + 1 + (n-2))/2 = ((n-3)a + n + 3)/2$,
- For any $2 < \ell < n - 1$ the following matrix is an optimal solution for P_ℓ.

row \ column	1	2	$\ldots$	$\ell-1$	ℓ	$\ell+1$	$\ldots$	$n-1$	n
1	0	1	$\ldots$	1	1	0	$\ldots$	0	0
2	1	$\ell-1$		0	0	0	$\ldots$	0	0
$\vdots$	$\vdots$		$\ddots$		$\vdots$		$\ldots$		$\vdots$
$\ell-1$	1	0		$\ell-1$	0	0	$\ldots$	0	0
ℓ	0	0	$\ldots$	0	$\ell-1$	1		0	0
$\ell+1$	0	0	$\ldots$	0	0	$\ell-1$	$\ddots$	0	0
$\vdots$	$\vdots$		$\ldots$		$\vdots$		$\ddots$	$\ddots$	
$n-1$	0	0	$\ldots$	0	0	0		$\ell-1$	1
n	1	0	$\ldots$	0	0	0	$\ldots$	0	$\ell-1$

and the value of this optimal solution is

$$\begin{aligned} F(\ell) &= \frac{1}{\ell}\left[(\ell-1)2 + (\ell-1)(\ell-1) + 1 + (n-(\ell+1))a + (n-\ell)(\ell-1) + 1\right] \\ &= \frac{1}{\ell}\left[(\ell-1)(n+1) + 2 + (n-\ell-1)a\right]. \end{aligned}$$

- For $\ell = n - 1$ we get $F(n-1) = n = F_T$. An optimal solution is

$$[x^{n-1}_{i,j}]_{n \times n} = \begin{pmatrix} 0 & 1 & \ldots & 1 \\ 1 & n-2 & & \\ \vdots & & \ddots & \\ 1 & & & n-2 \end{pmatrix}.$$

In Fig. 1 we illustrate the optimal tour for $\ell = n - 1$. Here, the numbers on the arcs correspond to the multiplicities of the arc usage in the optimal tour.

Thus, for this instance $\ell^0 \le n-1$. We finish the proof by showing that $\ell^0 \ge n-1$. Indeed, for $\ell = n-2$ and $n \ge 3$,

$$\begin{aligned} F(\ell-2) &= \frac{((n-2)-1)(n+1)+2+(n-(n-2)-1)a}{n-2} \\ &= \frac{(n-3)(n+1)+2+a}{n-2} \\ &= \frac{n^2-2n-1+a}{n-2} \\ &= n + \frac{a-1}{n-2} \\ &> n, \end{aligned}$$

which yields $\ell^0 > n-2$ as required. □

7. Optimal sequences in the general case

To prove the main result of the section we need the following lemma.

Lemma 2. *For $\ell \in \mathbb{N}$, any feasible solution x of P_ℓ can be decomposed into ℓ feasible solutions of the transportation problem T, i.e., there exist ℓ feasible solutions $y^1, y^2, \dots, y^\ell$ of the transportation problem T such that $x = y^1 + y^2 + \cdots + y^\ell$.*

Proof. Let B be the constraint matrix of the transportation problem T and let s be the right-hand-side vector of T. Consider the polytope $\mathcal{P} = \{y \in \mathbb{R}^{n\times n} : \ By = s,\ 0 \le y \le x\}$. By definition of x, $Bx = \ell s$, and therefore $\mathcal{P}$ is not empty since it contains at least the point $y = x/\ell$. Since B is totally unimodular, $\mathcal{P}$ is not empty, and s and x are integral vectors, we derive that polyhedron $\mathcal{P}$ is integral; see, e.g., Theorem 19.1 in [13]. Therefore, there exists an integral solution $y^1 \in \mathcal{P}$. By definition of the polytope, the point y^1 is a feasible solution of the transportation problem T and it holds that $y^1 \le x$. Therefore, $y' = x - y^1$ is a feasible solution of the transportation problem $T_{\ell-1}$. If $\ell > 1$, repeatedly applying the same argument to y' decomposes y' into a feasible solution y^2 of the transportation problem T and a feasible solution y'' of the transportation problem $T_{\ell-2}$. Following this recursion we construct ℓ feasible solutions of the transportation problem T such that their sum equals x as required. □

Using Lemma 2 and Observation 1, we now extend Theorem 1.

Theorem 7. *For any natural number $\ell \ge n-1$ the following holds:*

$$F(\ell+1) = \frac{\ell}{\ell+1}F(\ell) + \frac{1}{\ell+1}F_T, \tag{32}$$

and

$$F(\ell) = F(n-1)\frac{n-1}{\ell} + F_T\frac{\ell-n+1}{\ell}, \tag{33}$$

and the solution $x^\ell = x^{n-1} + (\ell-n+1)x^T$ is an optimal solution of the problem P_ℓ, where x^{n-1} and x^T are optimal solutions of P_{n-1} and T respectively.

Proof. Let $\ell \ge n-1$. Consider an optimal solution x of P_ℓ and let $y^1, y^2, \dots, y^\ell$ be the solutions to T forming the decomposition of x as in Lemma 2. Let $G_x, G_{y^1}, G_{y^2}, \dots G_{y^\ell}$ again be the graph representations of $x, y^1, y^2, \dots, y^\ell$ respectively.

Without loss of generality assume that G_{y^1} has the minimum number of strongly connected components over all graphs $G_{y^1}, G_{y^2}, \dots, G_{y^\ell}$. Notice that the number of strongly connected components in G_{y^1} is at most $n-1$. Let I^1 be a strongly connected component of G_{y^1}. Since x satisfies the subtour elimination constraints (4), the graph G_x is strongly connected and hence for some $z^1 \in \{y^2, y^3, \dots, y^\ell\}$ it holds that $z^1_{i^*,j^*} > 0$ for some $i^* \in I$ and $j^* \in \overline{I}$. Without loss of generality let y^2 be this solution. By Observation 1, the number of components in $G_{y^1+y^2}$ is strictly smaller than the number of components in G_{y^1}. Now if $U_{y^1+y^2} > 1$, let I^2 be one of its strongly connected components. Again since G_x is strongly connected, define $z^2 \in \{y^3, y^4, \ldots, y^\ell\}$ such that $z^2_{i^*,j^*} > 0$ for some $i^* \in I$ and $j^* \in \overline{I}$. Again, without loss of generality we assume y^3 to be this solution, and apply Observation 1 and repeat the same argument. Continuing these arguments leads to a strongly connected multigraph in k steps, where $k \leq U_{y^1} - 1 \leq n - 2$.

The strong connectivity of $G_{y^1+y^2+\cdots+y^{k+1}}$ implies that the solution $x' = y^1 + y^2 + \cdots + y^{n-1}$ of the transportation problem T_{n-1} satisfies the subtour elimination constraints (4). Therefore, x' is a feasible solution of P_{n-1} which implies that for all $\ell \geq n-1$ the following holds:

$$\begin{aligned}(n-1)F(n-1) + (\ell - n + 1)F_T &\leq cx' + (\ell - n + 1)F_T \\ &= \sum_{l=1}^{n-1} cy^l + (\ell - n + 1)F_T.\end{aligned}$$

Since y^l, $l \in \{1, 2, \ldots, \ell\}$, are feasible solutions to the transportation problem T and F_T is the optimal objective value to this problem, we have that

$$\sum_{l=1}^{n-1} cy^l + (\ell - n + 1)F_T \leq \sum_{l=1}^{\ell} cy^l = cx = \ell F(\ell).$$

Hence, we derive that

$$(n-1)F(n-1) + (\ell - n + 1)F_T \leq \ell F(\ell). \tag{34}$$

Letting $\ell^* = n - 1$, combining (24) and (34) yields (33). Finally, (32) follows by induction from (33) as in Theorem 3. □

Notice that Theorem 7 provides an approximation preserving approach for construction of a good solution for problem P_ℓ with multiplicity parameter $\ell \geq n - 1$.

Corollary 3. *Let $\mathcal{A}$ be a polynomial time δ-approximation algorithm for the problem P_{n-1} and $F^{\mathcal{A}}$ be the objective value of the solution provided by $\mathcal{A}$. Then for all $\ell \geq n-1$, there exists a polynomial time δ-approximation algorithm for P_ℓ.*

Proof. Let us recall that $\mathcal{A}$ is a δ-approximation algorithm if

$$\frac{F^{\mathcal{A}} - F(n-1)}{F(n-1)} \leq \delta.$$

Now, consider an arbitrary $\ell \geq n - 1$. By Theorem 7,

$$\begin{aligned}F(\ell) &= F(n-1)\frac{n-1}{\ell} + F_T\frac{\ell-n+1}{\ell} \\ &\geq \frac{F^{\mathcal{A}}}{1+\delta}\,\frac{n-1}{\ell} + F_T\frac{\ell-n+1}{\ell} \\ &> \frac{1}{1+\delta}\left(F^{\mathcal{A}}\frac{n-1}{\ell} + F_T\frac{\ell-n+1}{\ell}\right).\end{aligned}$$

Consider the solution $x^{APPX} = x^{\mathcal{A}} + (\ell - n + 1)x^T$ of P_ℓ, where $x^{\mathcal{A}}$ is a solution provided by $\mathcal{A}$. Clearly, x^{APPX} is a feasible solution for P_ℓ and its objective value equals

$$F^{APPX}(\ell) = F^{\mathcal{A}}\frac{n-1}{\ell} + F_T\frac{\ell-n+1}{\ell}. \tag{35}$$

Therefore

$$F(\ell) > \frac{1}{1+\delta}F^{APPX}(\ell)$$

or equivalently

$$\frac{F^{APPX}(\ell) - F(\ell)}{F(\ell)} < \delta,$$

which completes the proof. □

Moreover, one can easily see that the approach taken in this section is not only approximation preserving but also asymptotically optimal with respect to multiplicity parameter ℓ.

Corollary 4. *Let x^{n-1} be a feasible solution to P_{n-1}, $x^{\ell} = x^{n-1} + (\ell - n + 1)x^T$ for all $\ell \geq n - 1$, and $\widehat{F}(\ell)$ be the objective value of P_ℓ corresponding to the solution x^ℓ. Then*

$$\lim_{\ell \to +\infty} \widehat{F}(\ell) = F_T.$$

Proof. The statement straightforwardly follows from equality (33) and the fact that F_T is a lower bound on any feasible solution of any problem P_ℓ, $\ell \in \mathbb{N}$. □

Acknowledgements

The paper has benefitted considerably from the constructive comments of the referees.

References

[1] A. Agnetis, No-wait flowshop scheduling with large lot size, Annals of Operations Research 70 (1997) 415–438.
[2] T. Ahuja, F. Magnanti, D. Orlin, Network Flows, Wiley Interscience, San Francisco, 1994.
[3] J.J. Clifford, M.E. Posner, Parallel machine scheduling with high multiplicity, Mathematical Programming 89 (2001) 359–383.
[4] S.S. Cosmadakis, C.H. Papadimitriou, The traveling salesman problem with many visits to few cities, SIAM Journal on Computing 13 (1984) 99–108.
[5] H.A. Eiselt, M. Gendreau, G. Laporte, Arc routing problems, part i: The chinese postman problem, Operations Research 43 (1995) 231–243.
[6] M.R. Garey, D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, Freeman and Company, San Francisco, 1979.
[7] A. Grigoriev, High multiplicity scheduling problems, Doctoral Thesis, Maastricht University, Maastricht, Netherlands, 2003.
[8] D.S. Hochbaum, R. Shamir, Strongly polynomial algorithms for the high multiplicity scheduling problem, Operations Research 39 (1991) 648–653.
[9] P. Kanellakis, C.H. Papadimitriou, Flow-shop scheduling with limited temporary storage, Journal of the ACM 27 (1980) 533–549.
[10] J. Van de Klundert, A note on the high multiplicity traveling salesman problem, Working Paper, Maastricht University, Maastricht, Netherlands, 1995.
[11] H.W. Lenstra Jr., Integer programming with a fixed number of variables, Mathematics of Operations Research 8 (1983) 538–547.
[12] G.L. Nemhauser, L.A. Wolsey, Integer and Combinatorial Optimization, Wiley-Interscience, Chichester, 1988.
[13] A. Schrijver, Theory of Linear and Integer Programming, Wiley-Interscience, Chichester, 1986.
[14] J.A.A. Van der Veen, S. Zhang, Low-complexity algorithms for sequencing jobs with a fixed number of job-classes, Computers And Operations Research 23 (1996) 1059–1067.